

# The Parameter Identification of Least Absolute Deviation Based on an Improved Gravitational Search Algorithm

Baochang Xu<sup>1,\*</sup>, Hua Zhang<sup>1</sup>, Likun Yuan<sup>1</sup> and Jinshan Wang<sup>2</sup>

<sup>1</sup>Department of Automation, China University of Petroleum (Beijing), Beijing, 102400, China

<sup>2</sup>PetroChina Tarim Oilfield Company, Korla ,841000, China

\*E-mail:xbcyl@163.com

**Abstract.** The least squares (LS) identification algorithm is vulnerable to outliers and has large residual square when the measured data is mixed with impulse noise which obeys symmetrical alpha stable (*SaS*) distribution, so the least absolute deviation (LAD) is selected as the objective function to get better identification performance when impulse noise exists. And taking the non-difference of least absolute deviation into consideration, we adopt an improved gravitational search algorithm as optimal algorithm to search for optimal solution globally. Then the parameter identification method based on LAD objective function using an improved gravitational search algorithm (LAD-IGSA) is put forward creatively. Simulation results show that the LAD-IGSA method can restrain the influence of impulse noise effectively and achieve higher identification accuracy. Moreover, LAD-IGSA method presents better robustness and better identification accuracy than LP method with small data sets.

## 1. Introduction

LS estimation is relatively mature and perfect in theory and methods, which has been widely applied to time series system identification, state estimation, functional approximation, and many other research fields [1]. LS also has some shortcomings: when measurement noise exists, especially, when impulse noise exists, LS shows poorer optimization precision [2]. In order to minimize the sum of squares, it's inevitable to submit to these outliers [3]. The studies show that the LS method has better identification effect when the stochastic noise is normally distributed. But the LS recursive method may come along with ill-posed solutions because of matrix inversion. Therefore, LAD is selected as the objective function to solve the problem of large residual when the measured data is disturbed by the impulse noise [4]. However, LAD does not gain enough attention in estimation and research compared with LS until Charnes applied LAD to a specific management question that describing the deviation with the difference between two nonnegative numbers [5]. This outstanding research laid the foundation of LAD. LAD avoids the complexities of matrix inversion and decreases the effect of impulse noise interference. Many experiments have proved that LAD demonstrates a strong robustness compared with LS [6]. But the LAD criterion function is non-differentiable, so the traditional optimization method based on gradient information can not be adopted. So far many effective algorithms for optimizing the LAD criterion function had been proposed such as linear programming [7], maximum entropy method [8] and intelligent optimization algorithm [9]. Intelligent optimization algorithm for the model parameters identification is essentially a probabilistic search, which does not



require the gradient information of the objective function, and optimizes the objective function without requirement of continuity.

In this paper, the parameter identification based on the least absolute deviation and an improved gravitational search algorithm (IGSA) is firstly put forward. Gravitational search algorithm (GSA) is a novel meta-heuristic stochastic optimization algorithm inspired by the law of gravity and mass interactions [10]. The GSA method has been employed to filter modelling [11] and unit commitment of renewable energy sources [12] and slope stability analysis [13]. The course of the parameter identification is also to seek the minimum of an objective function. It is important to carry out the precise neighbourhood search in the late period of optimization for enhancing the identification precision.

## 2. The improved algorithm of GSA

### 2.1. GSA

In GSA, considering a system with  $N$  agents which are a collection of masses. Each agent is composed of four characteristics: position  $X_i$ , inertial mass  $M_{ii}$ , active gravitational mass  $M_{ai}$  and passive gravitational mass  $M_{pi}$ , where  $i$  represents the  $d^{th}$  agent and  $X_i = (x_i^1, \dots, x_i^d, \dots, x_i^n)$ ,  $i = 1, 2, \dots, N$ ,  $x_i^d$  is the  $d^{th}$  dimension value of agent  $i$ .

According to Newton's gravitational force that the force  $F_{ij}$  acts on agent  $i$  by agent  $j$  at time  $t$  is

$$F_{ij}^d(t) = G(t) \frac{M_{pi}(t) \times M_{aj}(t)}{R_{ij}(t) + \varepsilon} (x_j^d(t) - x_i^d(t)) \quad (1)$$

Where  $\varepsilon$  is a small constant and  $G(t)$  is gravitational constant at time  $t$  which is decreased with time gradually so as to control the search accuracy.

$$G(t) = G_0 e^{-\tau \frac{t}{t_{\max}}} \quad (2)$$

Where  $G = 100$ ,  $\tau = 20$  and  $t_{\max}$  is the maximum number of iterations.

Assuming that the total force acting on object  $i$  in the  $d^{th}$  dimension is

$$F_i^d(t) = \sum_{j \in Kbest, j \neq i} rand_j F_{ij}^d(t) \quad (3)$$

Where  $Kbest$  means the  $Kbest$  agents with the best fitness value and biggest mass,  $rand_j$  is a random constant in the interval  $[0, 1]$  as well. With the initial value  $Kbest = N$  at the beginning,  $Kbest$  is a function that decreases with time. Under this circumstance, all agents apply the force mutually at the beginning, and with time going on,  $Kbest$  is decreased linearly. Eventually, only one object is left to apply force to other agents at the end [14].

As a result, the acceleration on the basis of Newton Second Law is

$$a_i^d(t) = \frac{F_i^d(t)}{M_{ii}(t)} \quad (4)$$

Gravitational and inertial masses are updated by the following equations:

$$M_{ai} = M_{pi} = M_{ii} = M_i, i = 1, 2, \dots, N, m_i(t) = \frac{fit_i(t) - worst(t)}{best - worst(t)}, M_i(t) = \frac{m_i(t)}{\sum_{j=1}^N m_j(t)}, \quad (5)$$

Where  $fit_i(t)$  represent the fitness value of the agent  $i$  at time  $t$ . Then the speed and position of the agent  $i$  will be updated as follows in GSA:

$$\begin{aligned} v_i^d(t+1) &= rand_i * v_i^d(t) + a_i^d(t), \\ x_i^d(t+1) &= x_i^d(t) + v_i^d(t+1). \end{aligned} \quad (6)$$

Where  $rand_i$  is a uniform random variable between 0 and 1.

## 2.2. Main improved strategies of IGSA

**2.2.1. Decrease the movement velocity of agents.** The next velocity of an agent is considered as a fraction of its current velocity added to its acceleration. Then in GSA, the speed and position of the agent  $i$  will be updated as (6). But the current velocity is so large that the current position is far from the next position. Some optimal position may be located between the current position and the next position that leads to the local minimum. Therefore, in IGSA the velocity of agents is decreased in order to improve the iterative search capability:

$$\begin{aligned} v_i^d(t+1) &= rand_i * v_i^d(t) + a_i^d(t) \\ x_i^d(t+1) &= x_i^d(t) + \eta * v_i^d(t+1) \end{aligned} \quad (7)$$

Where  $\eta$  is the number between 0 and 1.

**2.2.2. Orbital change of poor agents' positions.** During the course of GSA searching, all agents gradually converge to a small local zone, which results in a low searching efficiency in the late period, so orbital change operation should be established in order to jump out of the local minimum. And orbital change operation is to enlarge or contract the position of poor agents at a certain probability (named Jump rate).  $rands * x_i$  is called orbital change radius. That is, if agents' positions converge to a smaller value, the orbital change radius will be smaller. The orbital change operation is good for jumping out of the local minimum and improving the convergence speed, and not making a big disturbance upon the global.

$$xm_i = x_i + rands * x_i, i=1,2,\dots,N \quad (8)$$

Where  $rands$  is the random number between -1 and 1.

**2.2.3. Updating optimal agent using trial-and-error method.** In IGSA, the optimal object  $x_{best}$  is updated using the trial-and-error method. In GSA, all current agents change at each step, if the optimal agent's fitness becomes bad, the next search will begin from a worse position. The optimal position of those historical search steps,  $Lbest$ , and its fitness  $Fbest$ , only play a role for comparison, rather than participate into each step of iterative search. In order to utilize the information of those historical search steps  $Lbest$ , after one iteration is implemented for the optimal agents, the search will continue to the next step in case the fitness turns better. On the contrary, the position of optimal object's position and fitness will be replaced by  $Lbest$  and  $Fbest$ .

**2.2.4. Further search of optimal agent position.** The GSA algorithm generally converges quickly in the early 70% iterations, and then the convergence speed becomes slow. In order to further intensify the optimal searching ability of the algorithm in the late period, the optimal agent is further optimized by coordinate descent method and turns the multivariable optimization problem into some single-variable sub-problems.

## 3. Simulation and analysis for algorithm performance

### 3.1. Model description

Considering the time series as below:

$$\begin{aligned} A(z^{-1})y(k) &= B(z^{-1})x(k) + v(k) \\ A(z^{-1}) &= 1 + 1.4z^{-1} + 0.45z^{-2} \\ B(z^{-1}) &= z^{-1} + 0.7z^{-2}. \end{aligned} \quad (9)$$

Where the input sequence  $u(k)$  is white Gaussian noise with variance  $\sigma_u^2 = 1$ . The noise sequence  $v(k)$  is white Gaussian noise sequence with variance  $\sigma_v^2 = 0.2$ . The objective function can be

constructed as  $J(\theta) = \sum_{k=1}^m |e(k)|$ , where  $e(k) = y(k) - \varphi^T(k)\theta$ . LAD-IGSA is adopted to find the optimal solution of the LAD objective function. The evaluation standard in the experiments is the relative error  $\delta$  of the parameters estimation.

The impulse noise in the experiment subjects to *SaS* distribution, the probability density function of the standard *SaS* distribution is

$$f_{\alpha}(x) = \begin{cases} \frac{1}{\pi x} \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k!} \Gamma(\alpha k + 1) |x|^{-\alpha k} \sin(k\alpha\pi/2), 0 < \alpha < 1 \\ \frac{1}{\pi x} \sum_{k=1}^{\infty} \frac{(-1)^k}{2k!} \Gamma[(2k+1)/\alpha] x^{2k}, 1 < \alpha \leq 2 \end{cases} \quad (10)$$

Where  $\Gamma(*)$  is gamma function,  $\alpha$  is characteristics exponent. The less  $\alpha$  is, the more probability the large amplitude sample of the random variable occurs, and the stronger the pulse strength is [15,16].

### 3.2. Simulation results under different situations

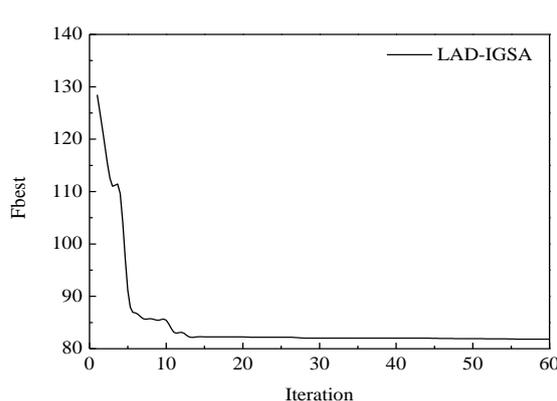
In order to guarantee the reliability of the simulation results, 100 times repeated independent experiments are carried out and the data length is  $L = 500$ .

#### Case 1: measured data with white noise only

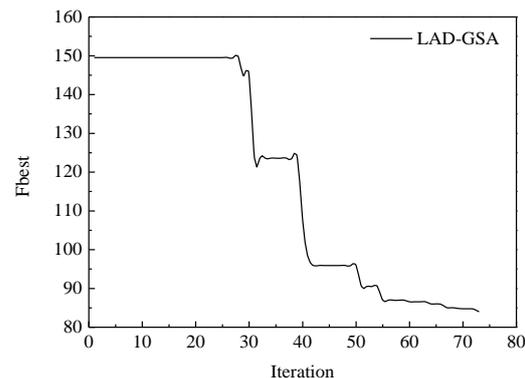
LAD-IGSA method, LAD-GSA method and LS method are adopted for parameter identification separately and the compared results are shown as Table 1, Figure 1 and Figure 2 are the curves of the optimal fitness  $F_{best}$  by LAD-IGSA method and LAD-GSA method.

**Table 1.** The identification results with white noise only

algorithm	$a_1$	$a_2$	$b_1$	$b_2$	$\delta$ (%)	iterations	time (s)
LAD-IGSA	-1.3883	-0.4403	1.0094	0.7001	0.94	60	2.5085
LAD-GSA	-1.4856	-0.5000	1.0314	0.8486	9.18	73	3.3522
LS	-1.4343	-0.4786	1.0097	0.7512	3.59	-	0.4742
true value	-1.4	-0.45	1.0	0.7	0	-	-



**Figure 1.** Curve of  $F_{best}$  by LAD-IGSA



**Figure 2.** Curve of  $F_{best}$  by LAD-GSA

$a_1$ ,  $a_2$ ,  $b_1$ ,  $b_2$  are estimated parameters. Table 1 indicates that the relative error of LAD-IGSA method is far less than that of LAD-GSA method and LS method. But LP method runs 0.4742s which presents higher estimation speed. Compared with Figure 1, the curve of  $F_{best}$  by LAD-GSA in Figure 2 has several parallel segments. That is to say, the objective function value does not always keep falling and falls into local optimal solution. Figure 1 presents that the improved strategies in LAD-IGSA method can improve the optimization accuracy and estimation speed greatly and ensures that the objective function value is optimized in the direction of minimization which contributes to jump out local optimal solution.

**Case 2: measured data with white noise and impulse noise which subjects to SaS distribution**

The measured data are contaminated with both white noise and impulse noise. Let the characteristics exponent  $\alpha$  of the impulse noise be 1.5, 1.2 and 0.9. Then the measured data are estimated by LAD-IGSA method, LAD-GSA method and LS method. The parameter identification results and the optimal fitness  $F_{best}$  of LAD-IGSA and LAD-GSA are shown in Table 2.

**Table 2.** The identification results when  $\alpha = 1.5$ 

algorithm	$a_1$	$a_2$	$b_1$	$b_2$	$\delta$ (%)	iterations	time (s)
LAD-IGSA	-1.3987	-0.4442	1.0041	0.6998	0.87	50	2.2535
LAD-GSA	-1.2957	-0.3711	1.0077	0.5641	9.83	98	4.8557
LS	-1.3656	-0.4272	0.9888	0.6607	3.04	-	0.0352
true value	-1.4	-0.45	1.0	0.7	0	-	-

**Table 3.** The identification results when  $\alpha = 1.2$ 

algorithm	$a_1$	$a_2$	$b_1$	$b_2$	$\delta$ (%)	iterations	time (s)
LAD-IGSA	-1.3976	-0.4486	1.0007	0.7186	0.67	40	1.9011
LAD-GSA	-1.2930	-0.3482	1.0049	0.5890	9.90	115	5.2288
LS	-1.4005	-0.4503	1.1630	0.8056	10.16	-	0.0351
true value	-1.4	-0.45	1.0	0.7	0	-	-

**Table 4.** The identification results when  $\alpha = 0.9$ 

algorithm	$a_1$	$a_2$	$b_1$	$b_2$	$\delta$ (%)	iterations	time (s)
LAD-IGSA	-1.3881	-0.4385	0.9975	0.7088	0.98	77	4.4196
LAD-GSA	-1.3265	-0.3783	1.0579	0.5497	9.97	116	11.4576
LS	-1.3960	-0.4434	0.5287	0.7966	25.17	-	0.0355
true value	-1.4	-0.45	1.0	0.7	0	-	-

From Table 2 ~ 4, the relative errors of the LAD-IGSA method are less than 1% which is much better than LAD-GSA method and LS method. With the characteristics exponent  $\alpha$  decreasing, the identification accuracy of LAD-IGSA method and LS method become worse. The relative error of LS method changes greatly and the identification accuracy is unstable. LS method even can not identify the parameters successfully when  $\alpha = 0.9$ . But LAD-IGSA method maintains high precision and the relative error converges to a small value under the interference of impulse with different intensities. So it can be summarized that LAD-IGSA method overcomes the impact of impulse noise successfully.

**Case 3: The influence of length of the sample data on identification accuracy**

We carry on simulation tests with LP method and LAD-IGSA method in the context of the measured data which are of different length. The compared results under different circumstances are as following in Table 5 ~Table 8.

**Table 5.** The identification results when  $L=10$ 

algorithm	$a_1$	$a_2$	$b_1$	$b_2$	$\delta$ (%)	iterations	time (s)
LAD-IGSA	-1.4077	-0.4433	1.1094	0.6726	5.92	10	0.1439
LP	-1.0432	-0.2391	0.9688	0.3179	29.54	-	0.1724
true value	-1.4	-0.45	1.0	0.7	0	-	-

**Table 6.** The identification results when  $L=50$ 

algorithm	$a_1$	$a_2$	$b_1$	$b_2$	$\delta$ (%)	iterations	time (s)
LAD-IGSA	-1.3340	-0.3678	0.9944	0.7048	5.53	17	0.2756
LP	-1.5751	-0.6168	1.0343	0.9341	17.7	-	0.1717
true value	-1.4	-0.45	1.0	0.7	0	-	-

**Table 7.** The identification results when L=100

algorithm	$a_1$	$a_2$	$b_1$	$b_2$	$\delta$ (%)	iterations	time (s)
LAD-IGSA	-1.3824	-0.4399	1.0506	0.6961	2.86	67	1.1820
LP	-1.5918	-0.6133	1.0378	0.9187	17.57	-	0.1862
true value	-1.4	-0.45	1.0	0.7	0	-	-

**Table 8.** The identification results when L=300

algorithm	$a_1$	$a_2$	$b_1$	$b_2$	$\delta$ (%)	iterations	time (s)
LAD-IGSA	-1.3834	-0.4273	1.0159	0.6879	1.81	42	1.2637
LP	-1.4890	-0.5176	1.0201	0.8129	8.38	-	0.2659
true value	-1.4	-0.45	1.0	0.7	0	-	-

Table 5. shows that the identification accuracy of LAD-IGSA method still can get an acceptable identification result and is higher than that of the LP method with L=10. The accuracy of the LP method indicates that LP can not successfully identify the estimation parameter effectively. From Table 6. and Table 7., we can conclude that the accuracy of LP method is improved with the length of sample data increasing such as L=50 and L=100, but the relative error is still unaccepted. On the contrary, the relative error of LAD-IGSA method is still maintained at 5% or less. Table 8 shows that the LP method can identify the parameters effectively when L=300. Obviously, LAD-IGSA method can identify the parameters effectively when the sample data is insufficient or the length is small. LAD-IGSA method shows better robustness than LP method in this situation.

#### 4. Conclusions

In this paper, we firstly expand the application of the LAD based on an improved gravitational search algorithm to the field of time series system identification. The improved strategies in LAD-IGSA enhance the optimization capabilities and decrease the convergence time. LAD-IGSA method can inhibit the influence of impulse noise which obeys  $s\alpha s$  distribution on the identification results and improve the robustness to a great extent due to the application of the LAD criterion. The improved gravitational search algorithm contributes to get a better optimization results when the length of measured data is small, which make the available of identification of time series system when the measured dada are not sufficient. The proposed algorithm is carried out easily and should be very valuable for those applications such as identification of complex chemical process.

#### References

- [1] Xie K G, Song Q K and Zhou J Q A Linear Regress Model Based on Least-Absolute Criteria *Journal of System Simulation* vol 14 no 2 pp 189-192 2002
- [2] Cao H R System Identification Based on the Least-Absolute Criteria *Control Theory and Applications* vol 28 no 7 pp 8-10 2009
- [3] Chen X R and Wu Y H A Problem about the Relationship between the Minimum L1-Norm Estimate and Linear Programming *Journal of Nanjing Institute of Technology* vol 18 no 5 pp 16-21 1988
- [4] Xu B C, Lin Z H, Zhang Y D and Xiao Y Y Extended Stochastic Gradient Identification Method for Hammerstein Model based on Approximate Least Absolute Deviation *Mathematical Problems in Engineering* vol 2016 Article ID 9548428 2016
- [5] Charnes A, Cooper W W and Ferguson R O Optimal estimation of executive compensation by linear programming *Management Science* vol 1 no 2 pp 138-151 1955
- [6] Wan S P Multi-sensor Information Fusion Method Based on Least Absolute Deviation Estimation *Computer Engineering* vol 36 no 2 pp 257-259 2010
- [7] Xiao J H and Yang W L Realization of Least-Absolute Regression Using Linear Programming Based on Matlab *Journal of Wuyi University(Natural Science Edition)* vol 20 no 1 pp 5-8 2006

- [8] Yao J K and Cui H Y Maximum Entropy Method for Nonlinear 11 Problem *Communication On Applied Mathematics and Computation* vol 15 no 1 pp 79-86 2001
- [9] Jing J and Wang J Application of genetic algorithm to nonlinear LAD regression *Hydro-Science and Engineering* no 2 pp 43-47 2008
- [10] Rashedi E, Nezamabadi-pour E and Saryazdi S GSA: a gravitational search algorithm *Information Sciences* vol 179 no 13 pp 2232-2248 2009
- [11] Rashedi E, Nezamabadi-pour E and Saryazdi S Filter modeling using gravitational search algorithm *Engineering Applications of Artificial Intelligence* vol 24 no 1 pp 117-122 2011
- [12] Shukla A and Singh S N Multi-Objective Unit Commitment With Renewable Energy Using GSA Algorithm *INAE Letters* vol 1 no 1 pp 21-27 2016
- [13] Jiang J G, Tan Y, Dong L M and Wang C Application of modified gravitational search algorithm in slope stability analysis *Chinese Journal of Geotechnical Engineering* vol 38 no 3 pp 419-425 2016
- [14] Bazaraa M S and Shetty C M Nonlinear Programming: Theory and Algorithms *John Wiley & Sons* pp 779-841 2005
- [15] Xu B C and Zhang Y Y An improved gravitational search algorithm for dynamic neural network identification *International Journal of Automation and Computing* vol 11 no 4 pp 434-440 2014
- [16] Coates M J and Kuruoglu E E Time-frequency based detection in impulsive noise environments using alpha-stable noise model *Digital Signal Processing* vol 82 no 3 pp 1917-1925 2002